FIG 1 Peter and the wolf: Score

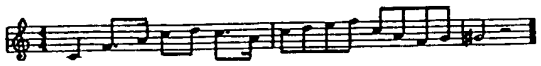

FIG 2 Peter and the wolf: Sound wave

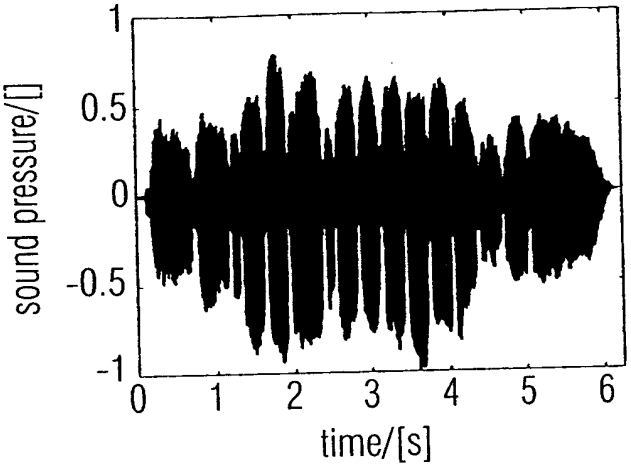


FIG 3 IHC cleft contents

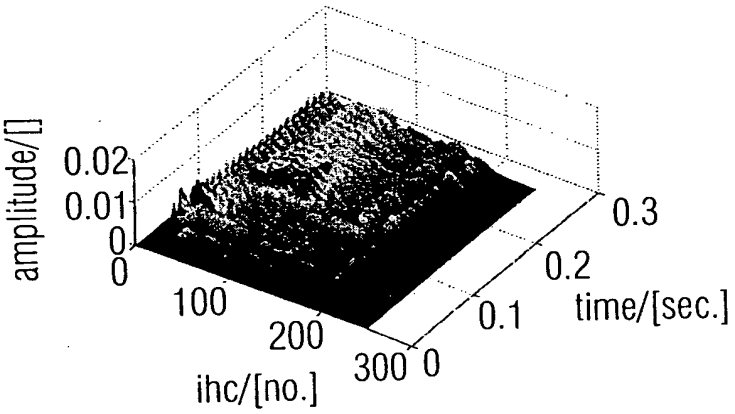

FIG 4. Cleft content of IHC #12

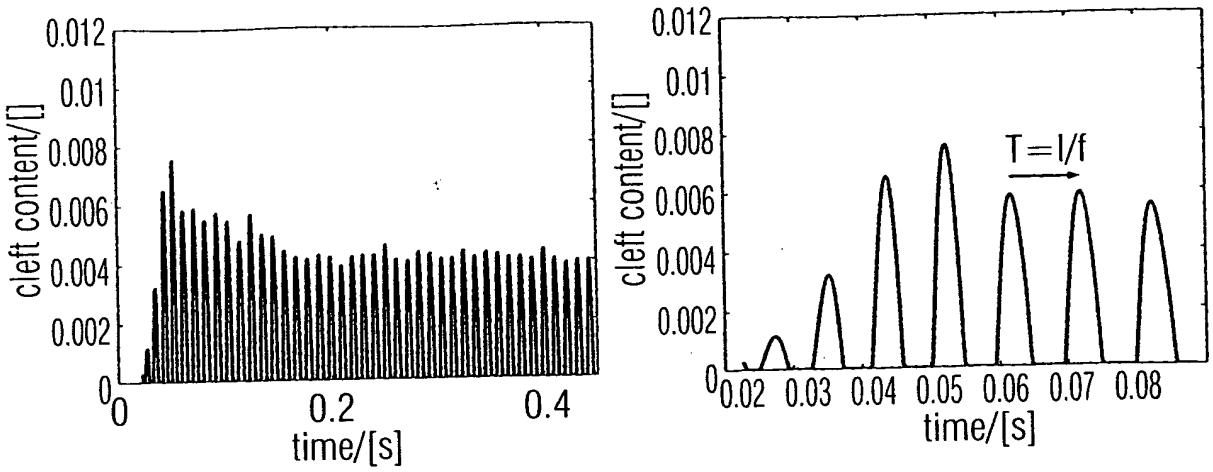


FIG 5. Cleft content of IHC #12

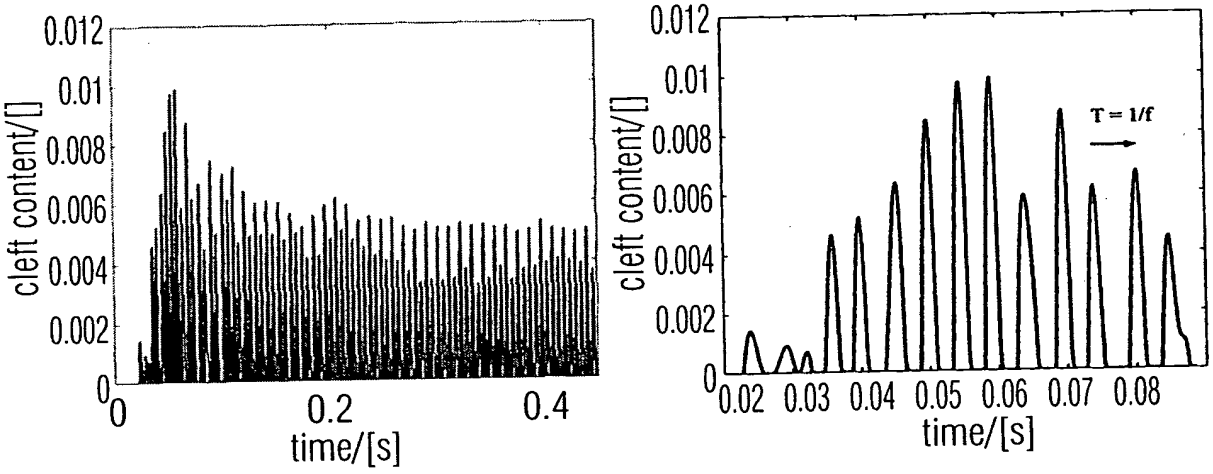


FIG 6. SACF histogram

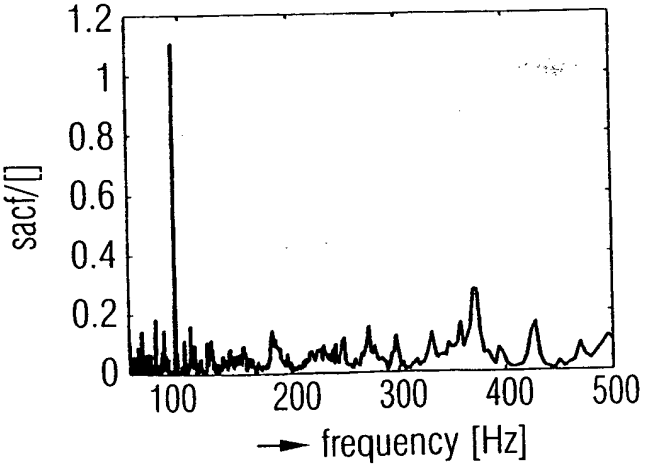

FIG 7: Initial pitch estimation vs. cleaned pitch trajectories

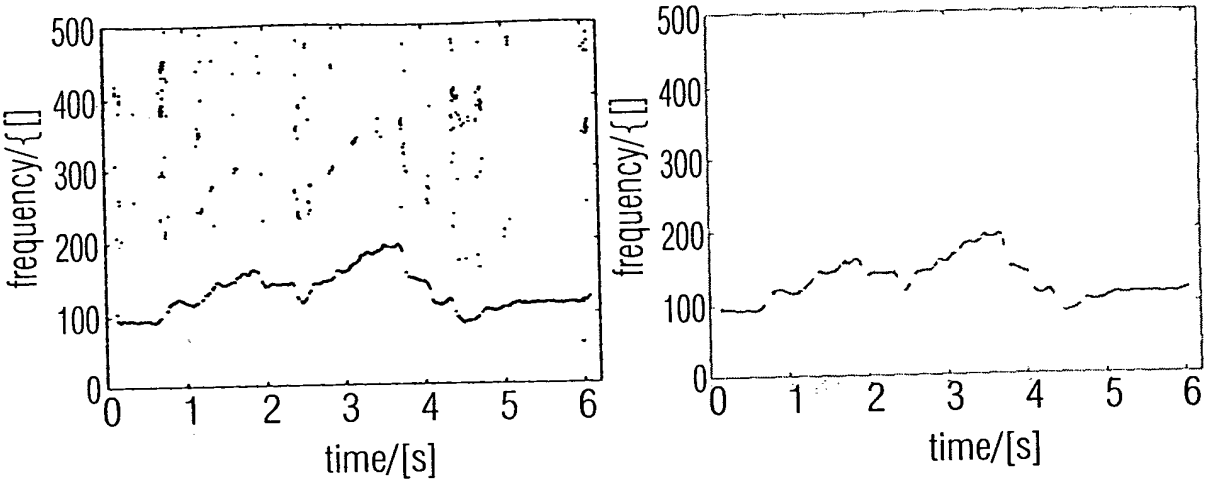


FIG 8: Envelopes of transmitter substance: 1st and 4th partial

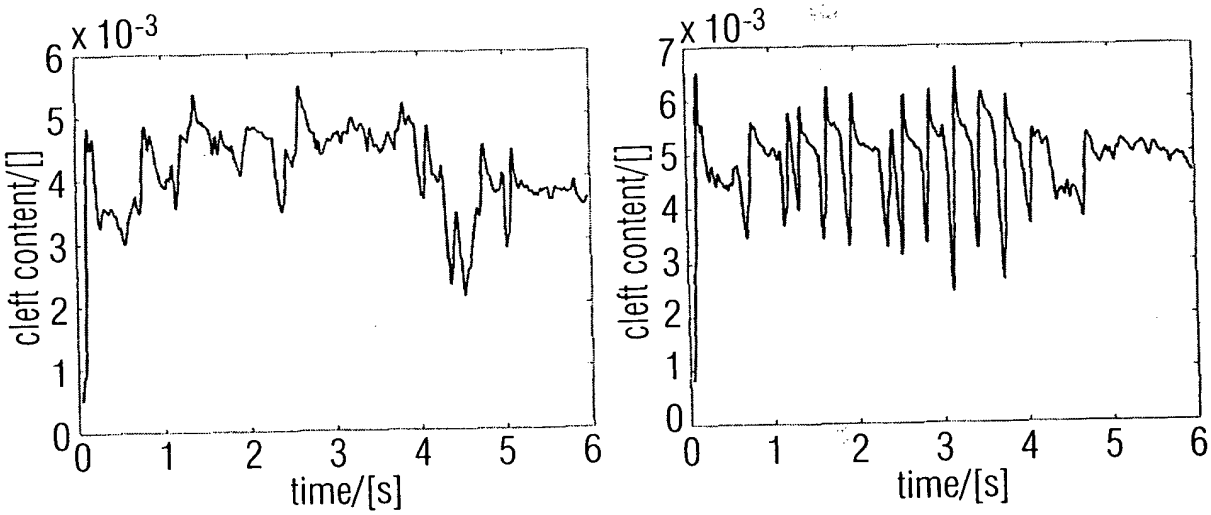

FIG 9: Onset map

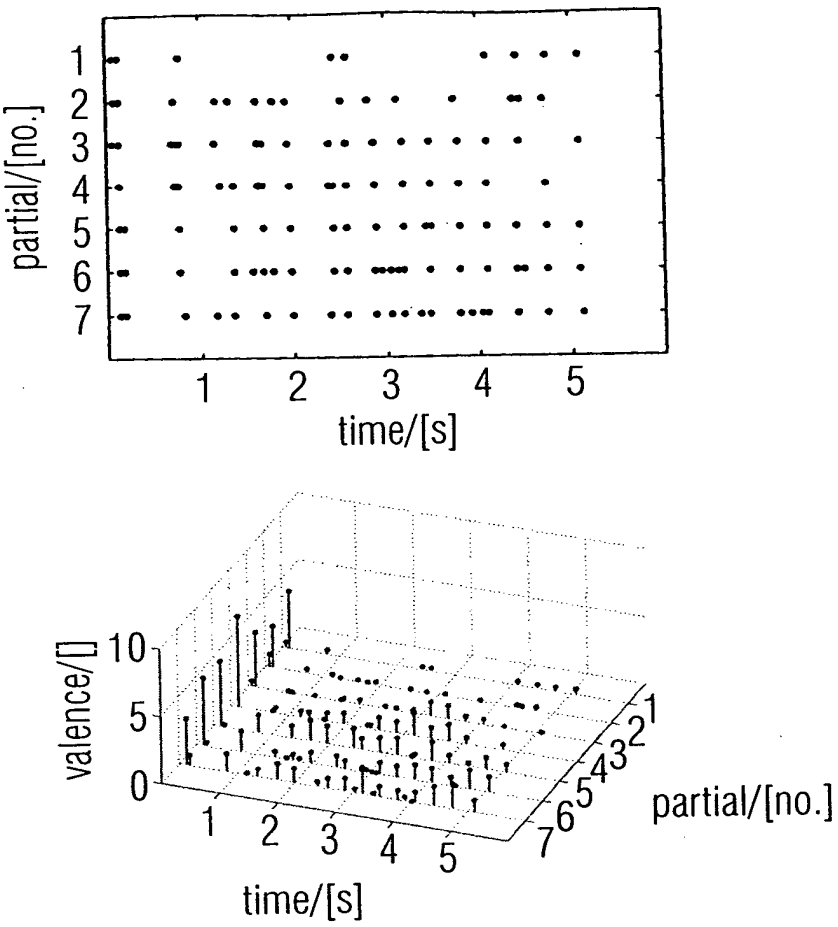


FIG 10: Onset masking

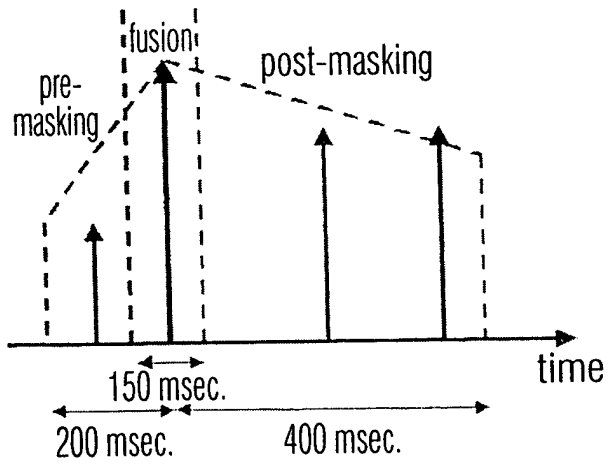


FIG 11: Onset histogram

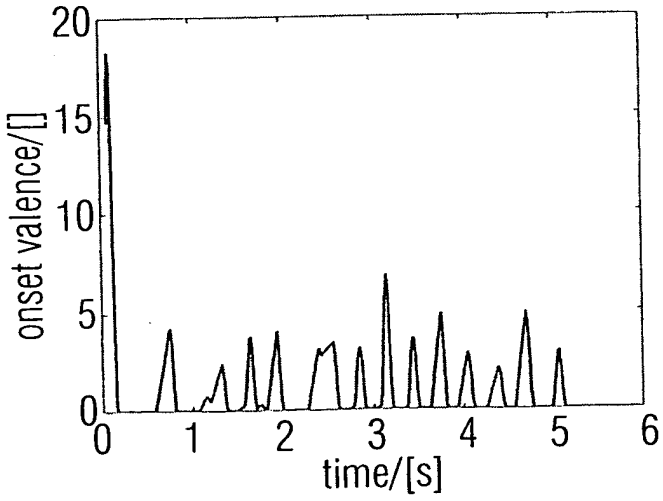

FIG 12: Pitch trajectories, segmented

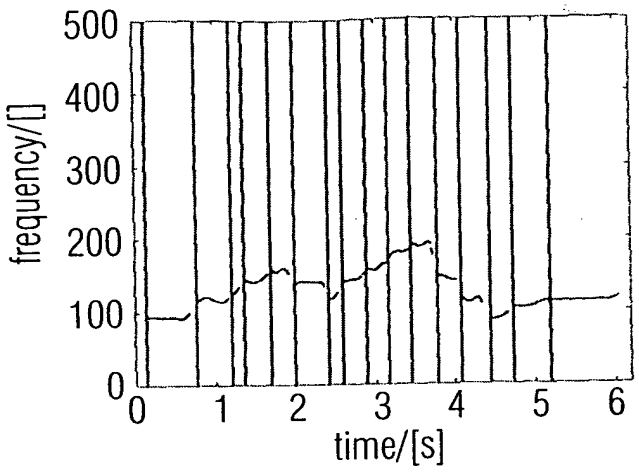


FIG 13: Feature values, clarinet Bb3

| partial [*no.*] | onset time [*s*] | frequency [*Hz*] | amplitude [] | IHC peak [*no.*] | IHC resonance [*no.*] |
|---|---|---|---|---|---|
| 1 | 0.187 | 236.4 | 0.00878 | 27 | 30 |
| 2 | 0.218 | 467.8 | 0.00555 | 48 | 14 |
| 3 | 0.192 | 703.5 | 0.00862 | 73 | 36 |
| 4 | 0.219 | 878.2 | 0.00735 | 94 | 2 |
| 5 | 0.201 | 1170.6 | 0.00733 | 102 | 12 |
| 6 | 0.202 | 1405.3 | 0.00776 | 113 | 12 |
| 7 | 0.206 | 1640.8 | 0.00783 | 123 | 16 |

FIG 14: Query-By-Humming: 1152 query inputs

| | | EarAnalyzer | Extreme | Hough |
|---|---|---|---|---|
| database 1 (200 pieces) | Position 1 | 82.8 % | 59.5 % | 39.9 % |
| | Top 10 | 92.5 % | 75.1 % | 56.9 % |
| database 2 (1024 pieces) | Position 1 | 78.5 % | 53.5 % | 32.0 % |
| | Top 10 | 88.9 % | 67.9 % | 42.9 % |

FIG 15: Query-By-Humming incl. GSM distortion: 1152 query inputs

| | | EarAnalyzer | Extreme | Hough |
|---|---|---|---|---|
| Fullrate | Position 1 | 79.3 % | 49.4 % | 42.0 % |
| | Top 10 | 88.9 % | 67.1 % | 56.3 % |
| Enhanced Fullrate | Position 1 | 80.6 % | 60.4 % | 45.0 % |
| | Top 10 | 91.2 % | 73.9 % | 62.7 % |
| Halfrate | Position 1 | 69.9 % | 48.4 % | 30.2 % |
| | Top 10 | 82.5 % | 65.4 % | 49.0 % |
| Original | Position 1 | 82.8 % | 59.5 % | 39.9 % |
| | Top 10 | 92.5 % | 75.1 % | 56.9 % |

FIG 16: Instrument recognition rates [total (clarinet|oboe|bassoon)]

| training vs. query | McGill [%] | Gdansk [%] | Fraunhofer [%] |
|---|---|---|---|
| McGill | 100 (100\|100\|100) | 80 ( 94\| 47\| 94) | 88 ( 97\| 67\| 97) |
| Gdansk | 85 ( 75\| 87\| 94) | 100 (100\|100\|100) | 87 ( 86\| 73\|100) |
| Fraunhofer | 81 ( 58\| 93\| 94) | 72 ( 64\| 60\| 94) | 100 (100\|100\|100) |
| McGill + Gdansk | 99 ( 97\|100\|100) | 100 (100\|100\|100) | 91 ( 92\| 83\| 97) |
| McGill + Fraunhofer | 100 (100\|100\|100) | 83 ( 92\| 53\|100) | 100 (100\|100\|100) |
| Gdansk + Fraunhofer | 88 ( 78\| 93\| 94) | 100 (100\|100\|100) | 100 (100\|100\|100) |

FIG 17: extended analog model by Baumgarte

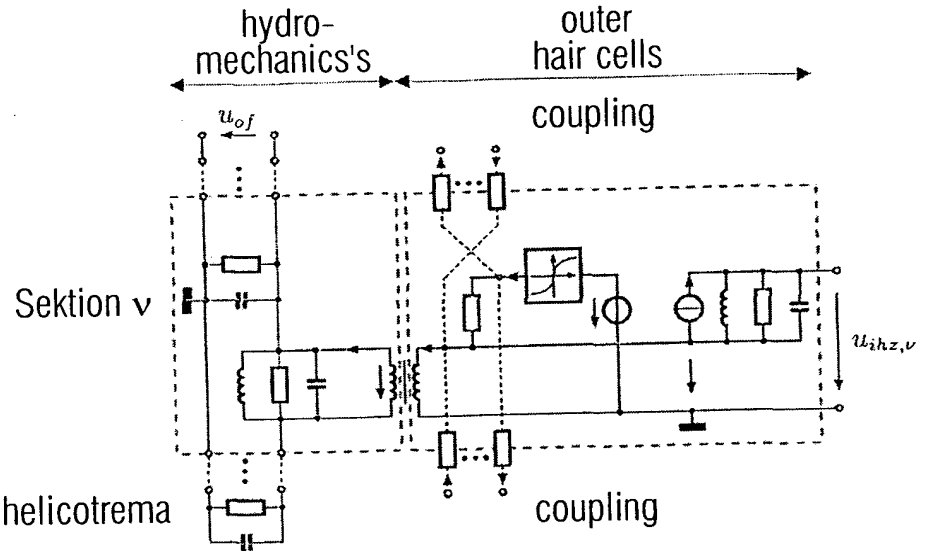


FIG 18: hair cell model by Meddis

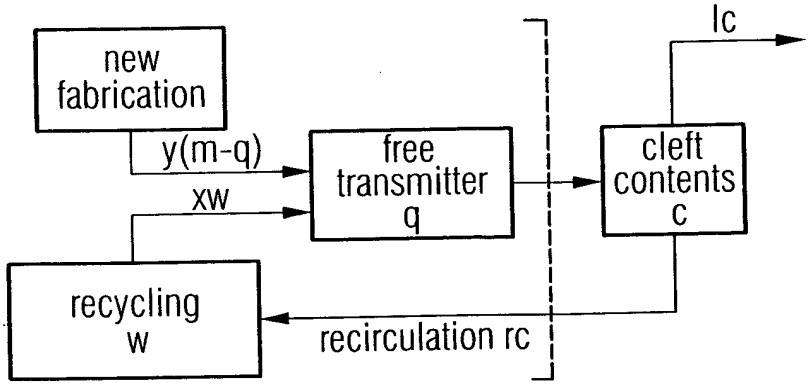

FIG 19: mathematical description of the model in Fig 18

$$k = \begin{cases} gdt \cdot \left(\frac{s+A}{s+A+B}\right) & : \quad s + A \geq 0 \\ 0 & : \quad s + A < 0. \end{cases}$$

$$\frac{dq}{dt} = y(m - q) + xw - kq,$$

$$\frac{dc}{dt} = kq - lc - rc,$$

$$\frac{dw}{dt} = rc - xw.$$

| | | |
|---|---|---|
| A | = | 10 |
| B | = | 3000 |
| x | = | 66.31 |
| g | = | 1000 |
| y | = | 5.05 |
| m | = | 1 |
| l | = | 2500 |
| r | = | 6580 |

FIG 20: crossection of the cochlea

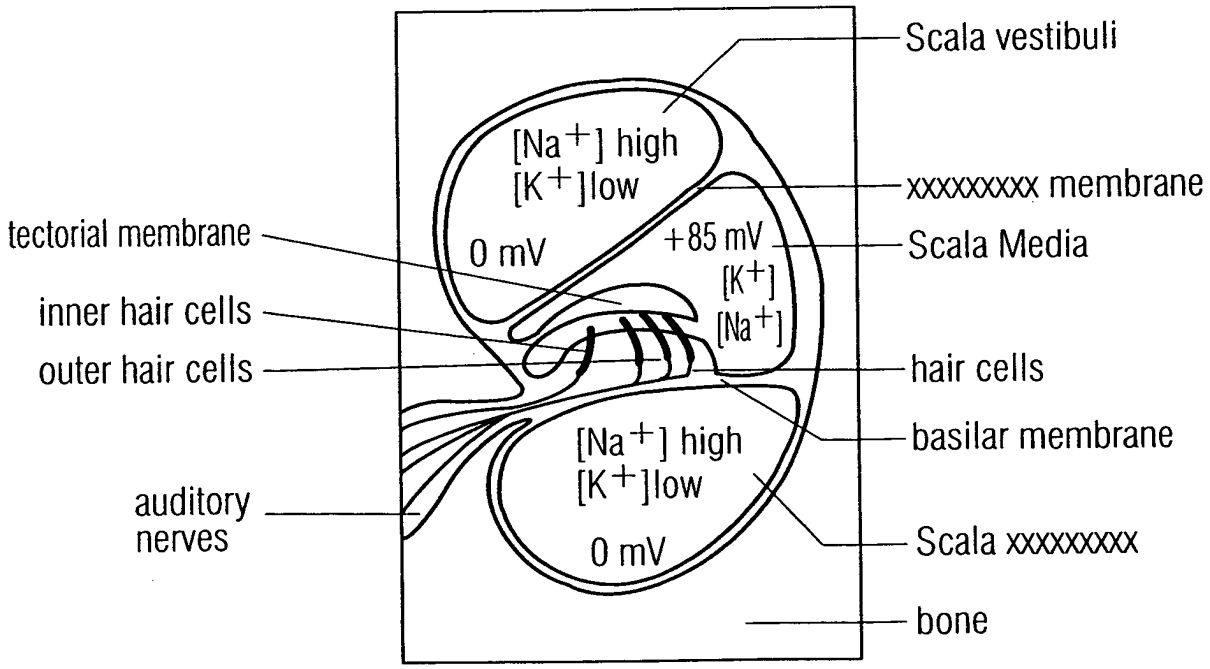

FIG 21

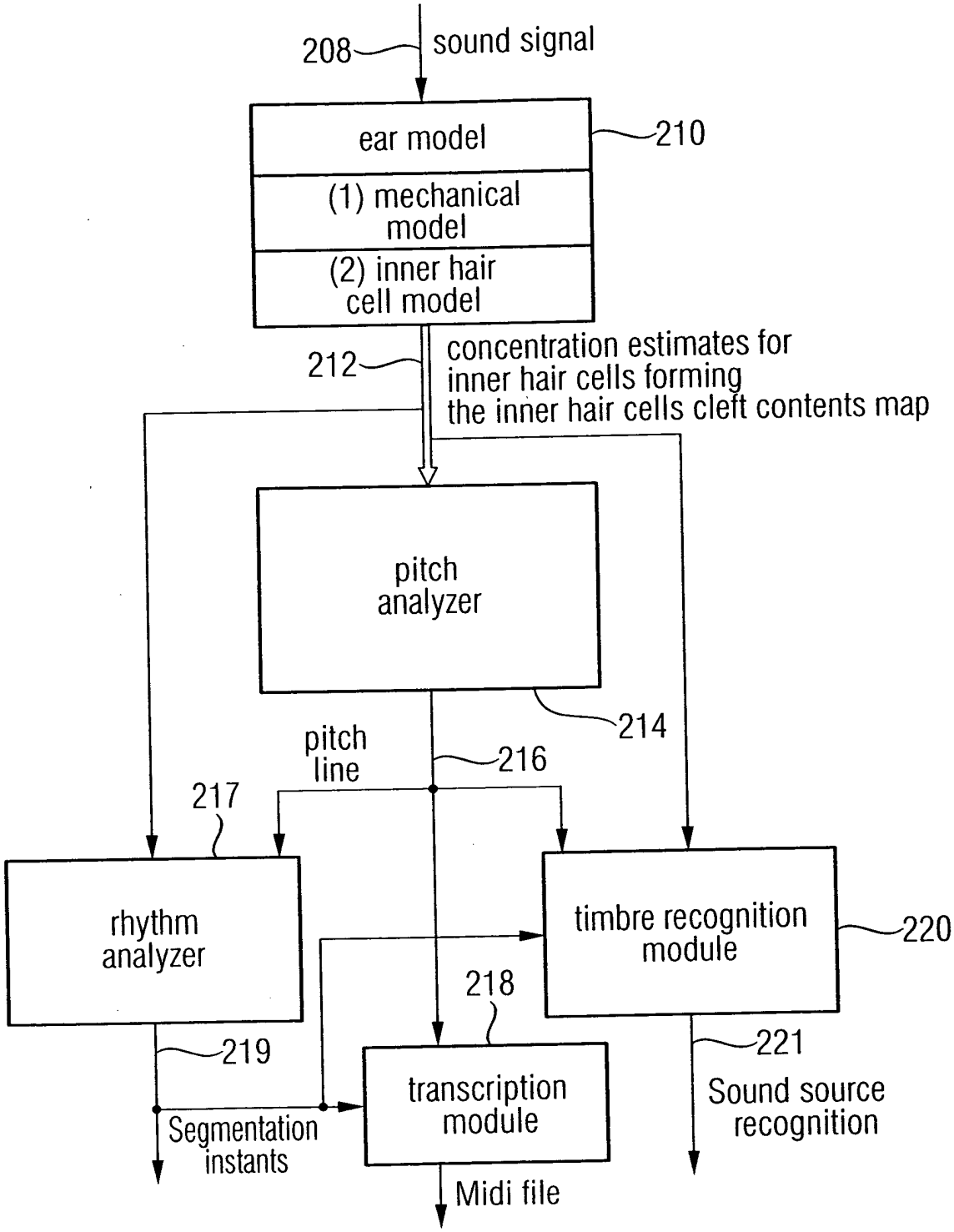
208
sound signal
ear model
210
(1) mechanical model
(2) inner hair cell model
212
concentration estimates for inner hair cells forming the inner hair cells cleft contents map
pitch analyzer
214
pitch line
216
217
rhythm analyzer
timbre recognition module
220
218
219
221
transcription module
Sound source recognition
Segmentation instants
Midi file

FIG 22 (pitch analyzer)

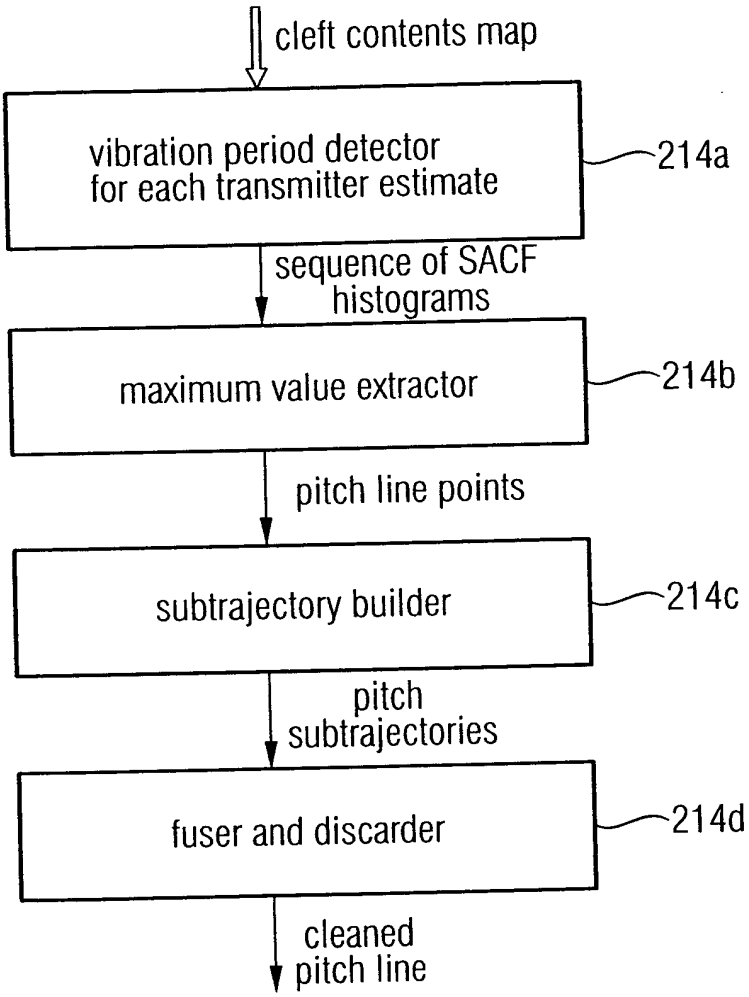

FIG 23 (rhythm analyzer)